AFRL-RI-RS-TR-2015-127

CORRECT-BY-CONSTRUCTION ATTACK-TOLERANT SYSTEMS

CORNELL UNIVERSITY

MAY 2015

FINAL TECHNICAL REPORT

APPROVED FOR PUBLIC RELEASE; DISTRIBUTION UNLIMITED

STINFO COPY

AIR FORCE RESEARCH LABORATORY
INFORMATION DIRECTORATE

■ **AIR FORCE MATERIEL COMMAND** ■ **UNITED STATES AIR FORCE** ■ **ROME, NY 13441**

NOTICE AND SIGNATURE PAGE

Using Government drawings, specifications, or other data included in this document for any purpose other than Government procurement does not in any way obligate the U.S. Government. The fact that the Government formulated or supplied the drawings, specifications, or other data does not license the holder or any other person or corporation; or convey any rights or permission to manufacture, use, or sell any patented invention that may relate to them.

This report is the result of contracted fundamental research deemed exempt from public affairs security and policy review in accordance with SAF/AQR memorandum dated 10 Dec 08 and AFRL/CA policy clarification memorandum dated 16 Jan 09. This report is available to the general public, including foreign nationals. Copies may be obtained from the Defense Technical Information Center (DTIC) (http://www.dtic.mil).

AFRL-RI-RS-TR-2015-127 HAS BEEN REVIEWED AND IS APPROVED FOR PUBLICATION IN ACCORDANCE WITH ASSIGNED DISTRIBUTION STATEMENT.

FOR THE DIRECTOR:

/ S /
ANTHONY MACERA
Work Unit Manager

/ S /
WARREN H. DEBANY, JR.
Technical Advisor, Information
Exploitation & Operations Division
Information Directorate

This report is published in the interest of scientific and technical information exchange, and its publication does not constitute the Government's approval or disapproval of its ideas or findings.

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

The public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Department of Defense, Washington Headquarters Services, Directorate for Information Operations and Reports (0704-0188), 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to any penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number.
PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ADDRESS.

1. REPORT DATE *(DD-MM-YYYY)*	**2. REPORT TYPE**	**3. DATES COVERED** *(From - To)*
MAY 2015	FINAL TECHNICAL REPORT	SEP 2010 – DEC 2014

4. TITLE AND SUBTITLE

CORRECT-BY-CONSTRUCTION ATTACK-TOLERANT SYSTEMS

5a. CONTRACT NUMBER FA8750-10-2-0238

5b. GRANT NUMBER N/A

5c. PROGRAM ELEMENT NUMBER 62303E

6. AUTHOR(S)

Robert Constable

5d. PROJECT NUMBER CRSH

5e. TASK NUMBER CO

5f. WORK UNIT NUMBER RN

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

Cornell University
Department of Computer Science
Bill & Melinda Gates Hall
Ithaca, NY 14853

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)

Defense Advanced Research Projects Agency
675 North Randolph St
Arlington, VA 22203-2114

Air Force Research Laboratory/RIGA
525 Brooks Road
Rome, NY 13441

10. SPONSOR/MONITOR'S ACRONYM(S) AFRL/RI

11. SPONSOR/MONITOR'S REPORT NUMBER AFRL-RI-RS-TR-2015-127

12. DISTRIBUTION AVAILABILITY STATEMENT

Approved for Public Release; Distribution Unlimited. This report is the result of contracted fundamental research deemed exempt from public affairs security and policy review in accordance with SAF/AQR memorandum dated 10 Dec 08 and AFRL/CA policy clarification memorandum dated 16 Jan 09.

13. SUPPLEMENTARY NOTES

14. ABSTRACT

The purpose of the Cornell research on Correct-by-construction attack-tolerant systems is to increase the capabilities of computer scientists and software engineers to build highly reliable and adaptive cloud based computing systems and demonstrate these capabilities on prototype systems. This research will also provide the Department of Defense with advanced systems and methods for cyber-warfare. The research on this AFRL/DARPA project created new technology to render cloud based computing more resistant to cyber-attack and more capable of monitoring system state. The new technology was deployed and tested in a distributed database. It could be deployed in critical DoD systems. In the course of this project, the Cornell team strengthened its formal tools and extended the science behind advanced formal methods. The project also educated exceptional graduate students in this new technology and the computer science behind it. The enriched science base and consequent advanced technology provide a firm basis for investigating other aspects of distributed systems, such as how to make use of execution monitoring to adapt to cyber-attacks that are based on invalidating the mathematical assumptions on which verification is based. The deployment of verified systems revealed to the Cornell team the limits of formal guarantees and opened new lines of investigation with the potential to combine detailed formal knowledge of system potential with evidence from anomalous behavior to react to potential attacks and discover remedies based on operating data.

15. SUBJECT TERMS

asynchronous distributed systems, attack-tolerance, fault tolerant systems, correct-by-construction protocols, formal methods, event logic, functional distributed processes, cyber security

16. SECURITY CLASSIFICATION OF:			**17. LIMITATION OF ABSTRACT**	**18. NUMBER OF PAGES**	**19a. NAME OF RESPONSIBLE PERSON**
					ANTHONY MACERA
a. REPORT	**b. ABSTRACT**	**c. THIS PAGE**			**19b. TELEPHONE NUMBER** *(Include area code)*
U	U	U	UU	11	**N/A**

Contents

ABSTRACT

The purpose of the Cornell research on Correct-by-construction attack-tolerant systems is to increase the capabilities of computer scientists and software engineers to build highly reliable and adaptive cloud based computing systems and demonstrate these capabilities on prototype systems. This research will also provide the Department of Defense with advanced systems and methods for cyber-warfare. The research on this AFRL/DARPA project created new technology to render cloud based computing more resistant to cyber-attack and more capable of monitoring system state.

The new technology was deployed and tested in a distributed database. It could be deployed in critical DoD systems. In the course of this project, the Cornell team strengthened its formal tools and extended the science behind advanced formal methods. The project also educated exceptional graduate students in this new technology and the computer science behind it.

The enriched science base and consequent advanced technology provide a firm basis for investigating other aspects of distributed systems, such as how to make use of execution monitoring to adapt to cyber-attacks that are based on invalidating the mathematical assumptions on which verification is based. The deployment of verified systems revealed to the Cornell team the limits of formal guarantees and opened new lines of investigation with the potential to combine detailed formal knowledge of system potential with evidence from anomalous behavior to react to potential attacks and discover remedies based on operating data.

1 SUMMARY

This report summarizes the main technical accomplishment of the project *on Correct-by-construction attack-tolerant systems*. The goal of our research before, during, and after the period covered by the AFRL/DARPA funding has been to defend distributed systems against cyber-attacks. The report presents our technical accomplishments at a high level with the goal of relating the underlying science to the broad strategic issues facing the DoD in waging modern cyber warfare.

We made considerable progress in providing new technology to defend distributed systems and cloud based information systems, and it is sufficiently mature that it could be deployed operationally. Meanwhile the threats to such information systems remain high because they are more widely used, and attackers have greater incentives to find and exploit weaknesses.

It is clear that the United States is in a cyber-war arms race in which the stakes are being raised every day. The cost of vulnerabilities is increasing and the cost of errors in systems is also high and growing. We see no approach to both defense and offense that will not depend on the nation's ability to use advanced computer science and substantial computing power to enhance the ability of our systems to detect attacks and automatically adapt and reconfigure to remain functional. The ability to safely adapt is critical in both defense and offense.

Some of our accomplishments were specific to the task of building a demonstration attack-tolerant system, ShadowDB [1], using novel protocol synthesis techniques. ShadowDB is a distributed database system whose key protocols were synthesized and were formally proven to satisfy precise formal specifications about their behavior. This demonstration system also embodied our new methods for using *code diversity* to render systems more attack-tolerant. By creating alternative protocols for key functions of the distributed database system, we were able to respond to attacks by changing the protocols on-the-fly. Our system was used at MIT's Lincoln Labs as part of the DARPA CRASH project. We published these results, and our methods have been studied and continued at other research centers. We believe that our results are sufficiently robust and effective that they could be deployed in DoD systems.

We were also able to achieve unanticipated goals developed during the course of the project for which we were awarded additional funds. This led to methods for rendering systems Byzantine fault tolerant and to ideas for monitoring distributed system behavior and responding to unusual events. We believe that these methods will allow the DoD to build more resilient and attack-tolerant systems. We are submitting new research proposals to investigate these ideas and approaches. Our core formal methods capabilities were considerably extended over the course of the CRASH project. It is likely that we are one of the few research groups in the world who could have achieved the goals we set and accomplished.

Our research also resulted in new concepts that have been studied by other groups and have been taught in graduate computer science courses. For example, our formal definition of a *functional distributed process* is being investigated in systems research groups, and our concept of an *event type* [2, 3] is the basis of a new class of formal specification for distributed computing tasks. This notion also has promise in building secure *cyber physical systems*, and with Cornell funding we have been able to install verified code into a robot operating system and test it.

We also developed new techniques for formally reasoning about the performance of our protocols. These techniques were essential to our results for significantly improving the speed of the synthesized protocols without compromising their correctness. We developed new methods of proving the correctness of code optimization techniques. These formal methods will be important in a variety of efforts to create high performance correct by construction code.

2 INTRODUCTION

The problem addressed by this research is widely known to the general public because cyber-attacks have seriously disrupted commercial financial transactions and are known to be threats to the power grid, the air traffic control system, and all aspects of the banking system. They are threats to our basic cyber infrastructure, and already in this time period we as a nation have come to understand that basically every corporation is in the software business and every war will involve cyber warfare. The basic problem for the defense is how to protect our critical infrastructure, and the basic problem for offense is how to make effective weapons. The science behind these weapons is not physics, it is computer science. On the other hand, the methods of mathematics are as vital in this domain as they were in the creation of nuclear weapons. But now there is a new way in which the mathematics is vital. We discuss the change in the science base next.

2.1 Advanced Computer Aided Verification

C*omputer aided mathematical verification and program synthesis* is at the core of the new technology for building and defending critical information systems. The protocol creation tasks in this project were largely accomplished with software systems called *proof assistants* or *provers* [4, 5]. These are the tools predicted by artificial intelligence research half a century ago [6, 7]. Without these proof assistants we could not have accomplished our goals, and the US could not adequately protect its critical systems. Over the course of this project we made the tools we use much stronger, principally the proof assistants and the special automated analysis tools that are integrated with them (such as model checkers, solvers, decision procedures, and tactics). *We advanced the state of correct by construction programming and extended it into the realm of distributed systems.*

The ***first step*** in building secure systems is to say precisely what they must accomplish given assumptions about their operating environment and say it with mathematics. For asynchronous distributed systems this has been a very hard problem for a long time. It has gradually been brought to the state in which we now know ways to solve it well enough that proof assistants can help. It has taken the quality of work recognized by at least seven Turing Award winners to reach this state (two by Americans, one a student of the PI). There will be more such awards as these methods are deployed. Our research has contributed significantly to solving the specification problem, and this DARPA project enabled us to extend and deepen that work. It has led to the concept of *functional distributed processes* and to *event types,* both central to our approach.

One key insight behind the work reported here has been to express distributed algorithms as abstractly as possible in the programming paradigm with the most rigorous mathematical foundation, the functional paradigm. Another key insight has been to extend the method of specifying programs that has been the most widely used and deeply understood, the use of *mathematically precise types*. We designed the event type formalism used in this project [2, 3] and built strong tools to support it such as the EventML executable specification language [8].

The ***second step*** in building secure systems is to check that their top level implementations meet the precise mathematical specifications that define the tasks the system must accomplish under the assumptions on the operating environment. After that step, the next ones are to show that as the high level code is translated to executable machine code, the correctness guarantees are preserved. This kind of task was solved by other teams in the CRASH project. This task is what computer scientists now call *building a verified software stack*. DARPA is supporting important work along these lines and is funding a mastery of this technology that can be transitioned to industry.

3 METHODS, ASSUMPTIONS, AND PROCEDURES

We focused on specifying and synthesizing the key distributed protocols used by industry to build services such as databases and file systems that achieve fault tolerance using coordinated replicas. *Consensus protocols* allow system builders to organize their systems as interacting fault tolerant *distributed state machines*. To achieve fault tolerance, the state machines are built from replicated processes on different hardware that communicate with each other to ensure that they agree on the next step to take and how to update their copy of the state of the machine.

One of the most widely used consensus protocols is called Multi-Paxos, and another simpler protocol that requires more replicas to guarantee fault tolerance is called Simple Consensus. For each style of consensus protocol, there are many ways to implement it. Thus it is possible to create many different implementations that vary in the data structures used and the precise way in which the code is organized. We had proposed to use our correct by construction synthesis

methods to build several provably correct variants of each of these two kinds of consensus protocol. Our synthesis technology made this methodology conceivable, and our goal was to make it real. We did precisely that.

In order to achieve acceptable performance from our synthesized protocols, it became necessary to optimize them. It is critical that the optimization techniques not compromise the correctness guarantees established for the synthesized code. To ensure this, we developed a *provably correct suite* of optimization transformations [9]. This work required us to extend our theory of functional processes to take into account code transformations. We created methods to formally establish the correctness of the optimized code. This is a major step forward in formal mathematical reasoning about the behavior of practical algorithms.

4 RESULTS AND DISCUSSION

We established the effectiveness of a new methodology for building highly reliable distributed protocols and integrating them into complete systems. We demonstrated mechanisms for protecting systems against a variety of attacks. This work also led to a substantial enhancement of our proof assistant and to a deeper understanding of the potential for designing and building entire distributed systems using proof assistants as programming assistants. We understand what needs to be done to accomplish this and how to partner with systems builders to demonstrate this methodology. We believe it is urgent to reach this next level.

5 CONCLUDING REMARKS

Given the central role of proof assistants in much of the work undertaken by the projects in the DARPA CRASH program, it is important to explore the possibility of achieving an order of magnitude improvement in the capabilities of proof assistants. It is clear from their steadily expanding use in a wide variety of applications from cyber security and formalized mathematics to their applications in education that these are tools of extraordinary potential for the military as well as industry and academia. They were created with substantial funding from the US, both from NSF and DoD, from the UK, from France through INRIA, and from Germany. Microsoft Research has also invested substantially in this area. We built the Nuprl proof assistant [4] used in this project, and contributed to the Coq proof assistant [5], and we know how to make proof assistants substantially more effective, but there has been insufficient incentive and funding to take proof assistants to a higher level and specifically explore their critical role in cyber warfare.

The United States has the capacity to undertake this effort at a larger scale than anyone else. It is primarily DoD funding that has enabled US research groups to build and maintain high capacity in this critical technology. Now there is much less funding for that core technology itself at

precisely the time when we have considerable evidence that it is essential to basic research in programming languages, distributed systems, and cyber security. It is also important to modern computer science education. The value in education has been well recognized in the EU to their advantage, and DoD funding has given US academics support in their efforts to expand the educational role of proof assistants. This role should be recognized, appreciated, and if possible expanded for its impact on the next generation of American computer scientists.

REFERENCES

1. Nicolas Schiper, Vincent Rahli, Robbert Van Renesse, Mark Bickford, and Robert L. Constable. Developing correctly replicated databases using formal tools. In *DSN 2014: The 44th Annual IEEE/IFIP International Conference on Dependable Systems and Networks*, pages 395-406. Atlanta, Georgia, June 2014.

2. Mark Bickford and Robert L. Constable. Formal foundations of computer security. In *Formal Logical Methods for System Security and Correctness*, volume 14, pages 29-52, 2008.

3. Mark Bickford, Robert Constable, and Vincent Rahli. Logic of Events, a framework to reason about distributed systems. In *Proceedings of the 2012 Languages for Distributed Algorithms (LADA) Workshop*, Philadelphia, Pennsylvania, January 2012.

4. Stuart Allen, Mark Bickford, Robert Constable, Richard Eaton, Christoph Kreitz, Lori Lorigo, and Evan Moran. Innovations in computational type theory using Nuprl. *Journal of Applied Logic*, 4(4): 428-469, 2006.

5. Yves Bertot and Pierre Castéran. *Interactive Theorem Proving and Program Development; Coq'Art: The Calculus of Inductive Constructions.* Texts in Theoretical Computer Science. Springer-Verlag, 2004.

6. J. McCarthy. Computer programs for checking mathematical proofs. In *Proceedings of the Symposium in Pure Math, Recursive Function Theory*, volume V, pages 219-228. AMS, Providence, RI, 1962.

7. J. McCarthy. A basis for a mathematical theory of computation. In P. Braffort and D. Hirschberg, editors, *Computer Programming and Formal Systems*, pages 33-70. North-Holland, Amsterdam, 1963.

8. Vincent Rahli. Interfacing with Proof Assistants for Domain Specfic Programming using EventML. Presented at *The 10th International Workshop on User Interfaces for Theorem Provers*, Bremen, Germany, July 2012.

9. Vincent Rahli, Mark Bickford, and Abhishek Anand. Formal program optimization in Nuprl using computational equivalence and partial types. In *The 4th Conference on Interactive Theorem Proving (ITP 2013)*, pages 261-278. Rennes, France, July 2013.

10. Vincent Rahli, Nicolas Schiper, Robbert van Renesse, Mark Bickford, and Robert Constable. A Diversified and Correct-by-Construction Broadcast Service. In *The 2nd International Workshop on Rigorous Protocol Engineering (WRiPE),* Austin, Texas, October 2012.